ENGLISH HOUSING SURVEY (EHS)

Housing Stock Dataset
User Guide

2012

Contents

Introduction to the English Housing Survey (EHS)
- **Reporting**
- **EHS datasets**
- **Overview of the EHS housing stock dataset**
- **EHS database**
- **Other sources of information**

Using the database - first steps
- **How the housing stock data has been produced**
- **Sample structure**
- **Datasets on the UK Data Archive (UKDA)**
- **Levels of data**
- **Types of data**

Key concepts and definitions

Locating information in the database
- **Interview (household) survey variables**
- **Rotating modules**

How to match files

How to apply grossing factors to obtain national estimates

How to deal with missing data
- **Missing data conventions**
- **Dealing with missing data in analysis**
- **Checking if results are significant**

Appendix A: Content and level of interview survey files
Appendix B: Content and level of physical survey files
Appendix C: Content of detailed derived variables

Introduction to the English Housing Survey (EHS)

1 The EHS is a national survey commissioned by the Department for Communities and Local Government (DCLG) that collects information about people's housing circumstances and the condition and energy efficiency of housing in England.

2 The survey has a complex multi-stage methodology consisting of 2 main elements: an initial interview survey of around 13,300 households with a follow up physical inspection of a sub-sample of approximately 6,200 of these dwellings, including vacant dwellings.

3 In 2011-12, a cost review of the EHS resulted in several changes to the survey. The main changes to the survey were:

- a reduction to the EHS sample size
- a reduction in the questionnaire content. This included the removal of the EHS from the Integrated Household Survey (IHS)
- the cessation of the desk based market valuation of sub-sampled properties

Information on these changes can be found in the document "Proposals for changes to the English Housing Survey".

4 The English Housing Survey comprises of:

- an initial interview with approximately 13,300 households a year. This is referred to as the 'full interview sample'. This dataset is made available on a single financial year basis as the EHS household dataset.
- A follow up physical inspection of around 6,200 of these respondents' homes to assess the condition and energy performance of the property. This is referred to as the 'dwelling sample'. This dataset is made available on a rolling two-year basis. The 'dwelling sample' comprises all cases where an occupied dwelling has an interview and, physical survey completed plus vacant dwellings with a physical survey only.

5 There is a range of information that will support anyone wishing to make use of the survey data to conduct their own analysis or wishing to understand in detail how the survey is run and managed. All users are encouraged to familiarise themselves with the technical background before undertaking any detailed work using the EHS results.

Reporting

6 There are 5 reports which are published following each survey year. They are:

a) **EHS Headline Report** – the EHS Headline report is usually published around February time. This short report presents preliminary headline findings including key indicators related to departmental housing policies in areas such as trends in tenure and household composition, overcrowding, housing costs, and the condition and energy performance of the stock. Analysis is at a national level from both the full and dwelling samples. This is followed by two separate annual reports, later in the year.

b) **Households Report** – the EHS Households Report is published in the summer. It is based on the full interview sample only and presents comprehensive analysis of housing trends across each of the sectors and for different household groups including changing tenure patterns, overcrowding and under-occupation, rents and mortgages. New topics are introduced on a rolling basis reflecting changing policy interests.

c) In 2011 a single Homes report was produced; this has been split into three separate reports for 2012:

i) **Profile of English Housing** – which covers type and age of dwellings, profile of new builds, amenities and services, accessibility and dwelling condition

ii) **Energy Efficiency of English Housing** – insulation, heating, SAP rating, carbon dioxide emissions, hard-to-treat homes, renewable energy

iii) **Fire and Fire Safety** – smoke alarms and other fire safety measures, fire hazards

The reports also published in the summer and are based primarily on the rolling two-year dwelling sample i.e. cases where a physical inspection has been undertaken. The results presented for '2012' cover the fieldwork period April 2011 to March 2013. The one-year full interview sample (all household interview cases) or two-year household subsample (cases where a physical inspection and household interview has been undertaken) are used on occasion for topics analysing questions from the interview survey.

These are available to download from the EHS pages on the DCLG website. The reports include an overview of the survey methodology with guidance about sampling errors and a glossary of terms.

EHS datasets

7 The public datasets that will be available to external users through the UK Data Archive reflect the above EHS reporting strategy. Two separate datasets are therefore available as follows:

a) **EHS Household Dataset** – as used for the EHS Households report. The dataset comprises the full interview data (plus associated derived variables) for all cases where an interview has been completed – 13,300 households per annum (approximately 17,000 per annum before the EHS cost review). Datasets are provided for single financial years together with annual weights. This dataset should be used for any analysis where only information from the household interview is required.

b) **EHS Housing Stock Dataset** – as used for the Profile of English Housing, Energy Efficiency of English Housing, Fire and Fire Safety reports. The dataset is available for all cases where a physical survey has been completed and for occupied cases comprises data from the household interview as well as data from the physical survey. For vacant properties only data for the physical survey is provided. The data is made available for a two-year rolling sample with the appropriate two-year weights. For example, the EHS Housing Stock Dataset is for '2012' covering the period April 2011 to March 2013. This dataset comprises of approximately 14000 cases (previously 16,000 prior to the EHS cost review). The Housing Stock Dataset should be used for any analysis requiring information relating to the physical characteristics and energy efficiency of the housing stock.

8 **This User Guide relates solely to the EHS Housing Stock Dataset.** Users who wish to access the EHS Household dataset should refer to the separate EHS Household User Guide that accompanies that dataset. DCLG have deposited the data in this way to provide greater continuity for previous users of the Survey of English Housing (SEH) and English House Condition Survey (EHCS) and so that discrete datasets can be made available in the most straightforward way.

9 Each year new data files are created, some are dropped and a number of new variables are introduced. This applies particularly to the primary data files and users will need to be particularly careful when using these files. Further information is available in the technical background section of the website.

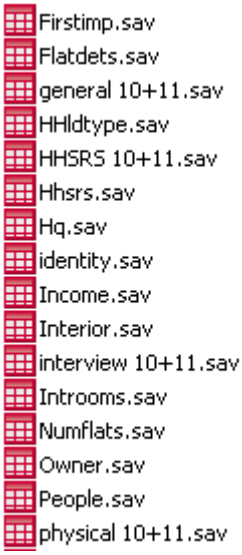
Overview of the EHS Housing Stock Dataset

10 The EHS Housing Stock data is available in SPSS, STATA and TAB format through the UKDA, as a download. To download or access the data you will need to register with the archive, though it is possible to

download descriptions of data and a general overview of the survey without registering.

EHS database

11 Once you have downloaded the EHS data in the preferred format, you will see a number of files within WinZip. The first lot are the **data files**, shown below in SPSS format.

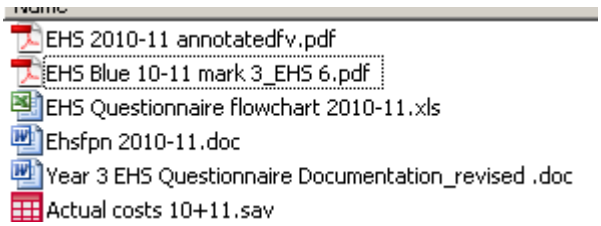


12 The EHS housing stock dataset consists of primary data files containing basic survey data collected by the surveyors and interviewers. These will be files such as firstimp.sav, interior.sav, dwelling.sav and numflats.sav etc

13 There are also a number of files containing 'derived' variables on the database. These are variables created by calculating with or recoding the basic survey data. These include energy performance 11+12.sav, equivalised income 11+12.sav and physical 11+12.sav

14 Users are encouraged wherever possible to make use of derived files for their analysis. This is because although the physical survey files have been checked for dimensional consistency and plausibility, they have not been through the rigorous editing, imputation and modelling that the derived variables have. It is advised that researchers using the primary data files to conduct analysis should be treating these files with more caution.

15 The second lot of files are the **documentation and reports**

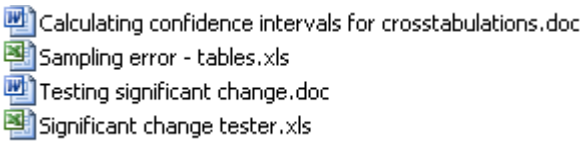

EHS 2010-11 annotatedfv.pdf
EHS Blue 10-11 mark 3_EHS 6.pdf
EHS Questionnaire flowchart 2010-11.xls
Ehsfpn 2010-11.doc
Year 3 EHS Questionnaire Documentation_revised .doc
Actual costs 10+11.sav

16 A number of documents are included on the EHS housing stock database to help users understand how the database is organised and which data files to use.

17 This includes the **interview survey documentation** which shows all the questions asked of households, the variable names and available response categories and showcards. Some new questions have been introduced and others have been dropped compared to previous years so the 2012-13 interview questionnaire provided indicates the new interview content. Certain topics are only collected on a rotating basis.

18 Also available on the dataset are the **physical survey documentation**. This includes a modified pdf version of the survey form to show the variable names alongside each question; surveyor instructions and the guidance that normally appear on the facing pages of the paper survey form.

19 The last lot of files on the zip drive are tools to help you with your analysis and reporting, including a quick way of checking whether your comparisons are statistically significant.



20 For 2012-13, a new version of the **data dictionary** has been produced to provide comprehensive details for all the variables on generalfs12.sav, general12.sav, interviewfs12.sav, interview12.sav and physical12.sav files. For information on the variables shown in the primary data files, please refer back to documentations delivered in previous years.

21 **Fuel poverty documentation** – a small number of variables relating to fuel poverty are included on the database. The link below provides further information relating to Energy Efficiency and Fuel Poverty. https://www.gov.uk/government/organisations/department-of-energy-climate-change/series/fuel-poverty-statistics

Other sources of information

22 Further information about the way in which the survey is organised is available on the EHS pages on the DCLG website.

23 As mentioned earlier in this note, all users are encouraged to familiarise themselves with the technical background known as the EHS Technical Report before undertaking any detailed work using the EHS data. These provide details of how the fieldwork is organised, how the sample is drawn and the results grossed each year, estimates of sampling and measurement error together with a full glossary of terms and details of how derived measures such as energy efficiency and income are created.

24 Alongside the EHS reports and their accompanying tables and figures, a series of **web tables**, on topics not covered in the headline or annual reports are also updated each year and published on the EHS pages on the DCLG website.

Using the database - first steps

25 To use the dataset correctly you must first ensure that you familiarise yourself with and be aware of the following:

- How the dataset has been produced
- What information is published
- The key concepts and definitions used in the survey
- How to locate the information you want in the database (find both survey and computed data by variable and file names)
- How to match files
- How to apply grossing factors to obtain national estimates
- How to deal with missing data
- How to deal with sampling and survey error

Each of these is described briefly in the sections below but please note that this guide is not intended to be a comprehensive source of information. Reference is made to the more detailed explanations and guidance contained in other files, publications and documents which are either at the UKDA or on the DCLG website.

The guide assumes knowledge and experience of using SPSS.

How the housing stock dataset has been produced

26 The EHS is a set of interrelated data from household and physical surveys based on a *sample of dwellings*.

27 The following describes the components of the whole EHS:

- A questionnaire-based survey of households living in the sampled dwellings which provides information on household characteristics, resources, housing costs and views of respondents about their home and neighbourhood. This is known as the **interview survey and is based on households.**
- A physical assessment of a sub-set of the sampled dwellings where permission/access was available or which were vacant at the time the interviewer called. This provides information on dwelling characteristics and a professional assessment of dwelling and neighbourhood conditions. This is known as the **physical survey and is based on dwellings.**

28 More detail on how the two surveys are conducted can be found on the DCLG website and in the Technical Appendices to the Headline and Annual Reports.

Sample structure

Annual sample size	25,286
Frequency of fieldwork	2 months per quarter
Number of months of interview fieldwork	8
Coverage	England only
Clustering	No
Design description	Two-stage sample of addresses
Achieved response rate	59%
Household interviews achieved	13,652
Paired cases achieved	6,058 (an additional 246 productive vacant cases were achieved)

29 In order to achieve the number of interviews and surveys described above, an initial sample of 36,000 addresses was drawn from the Postcode Address File (PAF). These addresses were drawn as a systematic two-stage random sample from the Royal Mail's Small User PAF.

30 For the first stage of sampling, 1,808 geographic 'merged areas' were generated by combining neighbouring Lower Layer Super Output Areas (LSOAs), so that each 'merged area' contained about 12,500 addresses. Half (904) of these 'merged areas' were randomly selected (as a systematic stratified sample) for the 2012-13 sample, with the remaining half then allocated for the 2013-14 sample.

The second stage of sampling involved selecting a stratified systematic random sample of addresses within the 'merged areas' selected for the 2012-13 sample.

31 The advantage of this two-stage approach is that it reduces the fieldwork area to half the country so interviewer and surveyor travel time / costs are significantly reduced. The disadvantage is that for any single survey year the survey is partially clustered, which will result in a small loss in statistical efficiency. However, when analysing any two years of survey data the combined sample is entirely unclustered.

For each sampled address, the predominant tenure within the postcode that contained that address was identified and attached to the record.

Predominant tenure was identified using Experian's Residata classifications; addresses were then grouped into strata and sub-sampled at the rates of 63.8% for owner occupied, 67.7% for private rented and 100% for social rented. As a result, 25,286 addresses were issued to interviewers.

32 Full details of sample methodology, and the changes implemented in 2011-12 can be found in the EHS Technical Report on the EHS pages on the DCLG website

Datasets at the UKDA

33 The EHS is a National Statistic. This means the statistics meet user needs, are produced, managed and disseminated to high standards, and managed impartially and objectively in the public interest.

34 All data files in the EHS housing stock dataset contain two years' worth of data. The following two-year datasets are available at the UKDA.

EHS housing stock dataset	Annual data contents	Mid point
EHS2008	**2007-08 and 2008-09**	**April 2008**
EHS2009	**2008-09 and 2009-10**	**April 2009**
EHS2010	**2009-10 and 2010-11**	**April 2010**
EHS2011	**2010-11 and 2011-12**	**April 2011**
EHS2012	**2011-12 and 2012-13**	**April 2012**

The number of (unweighted) cases in the housing stock dataset each year is as follows:

EHS housing stock dataset	Unweighted cases, dwellings	Unweighted cases, households
EHS2008	**16,150**	**15,523**
EHS2009	**16,150**	**15,512**
EHS2010	**16,670**	**16,047**
EHS2011	**14,951**	**14,386**
EHS2012	**12,763**	**12,269**

Levels of data

35 Most of the data is at household or dwelling level. However, some of the primary data files contain data at sub-household and sub-dwelling level. These are indicated for the interview survey at Appendix A and for the Physical Survey at Appendix B. All the files containing derived variables are at dwelling or household level.

Types of data

36 **Primary survey data** – primary data collected in the field is often very specific and detailed and in many cases is only collected in order to provide the building blocks for computing more useful pieces of information – in particular the household composition and income measures used in the analysis of the survey. There will also be a small amount of missing data in the primary data collected in the field.

In many cases omissions or inconsistencies in the base variables will have been identified in the course of producing derived variables for final analysis. **Users are recommended wherever possible to make use of the final derived variables rather than the raw data from the primary files as these are likely to contain more complete**

information on key topics such as rent, mortgage payments and income.

37 **Derived/computed data** – while the information that is collected in the field provides all the base variables needed, a considerable amount of further processing and modelling is needed to fully validate the results and produce secondary variables needed for detailed analysis. Producing derived variables often involves reconciliation of conflicting information from different parts of the survey (e.g. tenure) and/or the imputation of missing data where this has been possible from other data collected in the survey.

38 It is important to note that changes introduced to some of the EHS modules will have a significant impact on any time series analysis and users should familiarise themselves with the relevant EHS Technical Advice Notes before drawing conclusions about changes over time. Revised variable names have been used wherever there has been a major break in methodology year on year to minimise the risk of data being misinterpreted.

39 **Missing data** – as with all surveys there will be some level of missing data on the primary data files where information was not available or for example, a respondent refused to answer specific questions. Most of the derived variables for the dwelling sample however have no missing data, as missing values have been imputed based on cases with similar characteristics. Where imputation has taken place there are accompanying flag variables to indicate the level of imputation. Imputation is not always achievable so for some of the key derived variables there will be missing data which you may wish to distribute proportionately using the tool we have provided within the stock dataset. The method of data imputation is described more fully in the EHS Technical Report.

40 Alongside the annual data files, the UK Data Archive holds supporting documentation such as questionnaires, specifications for derived variables and show cards.

41 Several variables in the 2011 datasets have been top and/or bottom coded. Data has also been cleaned of all identifiable variables to maintain the confidentiality of respondents. Some response categories may also have been condensed for disclosure control reasons.

Key concepts and definitions[1]

42 The Housing stock dataset relates to both dwellings and households. It is therefore important to understand how these terms are defined and applied in the survey.

[1] A full glossary of terms used can be found in the EHS Headline and Annual Reports

43 **Dwelling** – a unit of accommodation which may comprise one or more household spaces (a household space is the accommodation used or available for use by an individual household). A dwelling may be classified as shared or unshared. A dwelling is shared if:

- the household spaces it contains are 'part of a converted or shared house', or
- not all of the rooms (including kitchen, bathroom and toilet, if any) are behind a door that only that household can use, and
- there is at least one other such household space at the same address with which it can be combined to form the shared dwelling.

Dwellings that do not meet these conditions are unshared dwellings.

The EHS definition of dwelling is consistent with the Census 2011.

44 **Household** – one person living alone, or a group of people (not necessarily related) living at the same address who share cooking facilities and a living room or sitting room or dining area. The EHS definition of household is consistent with the Census 2011.

Locating information in the database

Interview (household) survey variables

45 The topics covered in the interview survey questionnaire are illustrated in Figures 1 and 2 below, and includes a series of demographic and housing questions.

46 The demographic questions used to represent the Integrated Household Survey (IHS) core questions which were asked on the EHS until 2010-11. Although the EHS no longer forms part of the IHS, the questions are still asked on the 2012-13 EHS, due to their relevance to the EHS. These are listed in Figure 1.

47 The second part of the questionnaire is devoted to housing topics. For example there is a module of questions on mortgages (routed to households in the owner-occupied tenure) and other modules on rents, adaptations for disability, and energy efficiency. These housing topics are summarised in Figure 2.

Please note that wherever variable names have remained unchanged from the former SEH or EHCS – i.e. when a question has been taken directly from one of the former surveys with the same routing, question wording and response categories – the variable name will be unchanged. Where there has been a small change – e.g. a change to a response category or a slight change to the routing – then a number will have been added to the variable name to indicate this is a variant

of the original variable e.g. Vnllrd2 – (From whom did the household rent the property?), previously this was Vnllrd.

Rotating modules

48 Because of pressure on the size of the interview survey, the EHS also uses 'rotating modules' to bring questions in and out of the survey to meet users' demands and policy needs. The EHS generally adapts a two years in/two years out cycle depending on policy interest. Full details are given on the interview survey questionnaire.

49 Due to the nature of the rotating modules, not all topics will be covered on the housing stock dataset. This is because the housing stock requires 2 full years' of data.
E.g. the adaptation module was asked in 2009 and 2010, but not in 2011. The housing stock dataset for 2011 (combined years 2010 and 2011) will not contain the adaptation module as there is only 1 year's worth of data. The housing stock dataset for 2010 (combined years 2009 and 2010) will contain it as there are 2 years' worth of data.

50 The precise set of files and data provided each year will therefore vary reflecting the changing content of the interview survey. Details of the annual changes are available in the EHS Technical Report.

51 The content of the physical survey remains largely unaltered year on year. Details of the annual changes are available on the website.

Multi-coded variables

52 The EHS contains a number of multi-coded questions. These are questions where households could give more than one answer; for example 'adaptations for disability' or 'source of funding for buying a property'.

In the EHS household dataset, all multi-coded questions have been stored as a series of binary variables, with a separate variable for each answer option. It is expected that users may find this approach facilitates easier analysis. The following example (Example A) has been taken from the EHS questionnaire documentation:

Example A

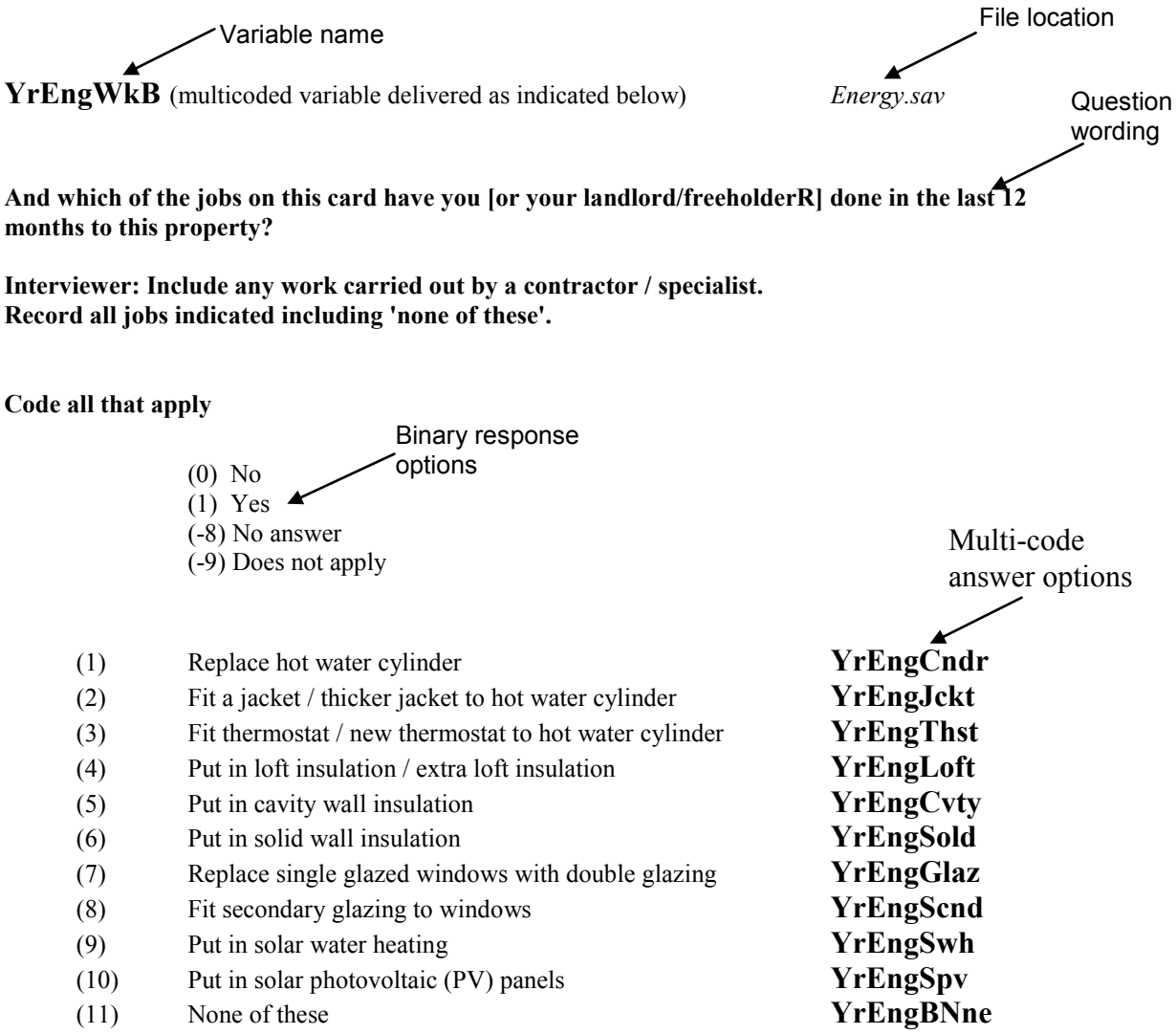

Variable name
File location
YrEngWkB (multicoded variable delivered as indicated below)
Energy.sav
Question wording
And which of the jobs on this card have you [or your landlord/freeholderR] done in the last 12 months to this property?
Interviewer: Include any work carried out by a contractor / specialist.
Record all jobs indicated including 'none of these'.
Code all that apply
Binary response options
(0) No
(1) Yes
(-8) No answer
(-9) Does not apply
Multi-code answer options
(1) Replace hot water cylinder YrEngCndr
(2) Fit a jacket / thicker jacket to hot water cylinder YrEngJckt
(3) Fit thermostat / new thermostat to hot water cylinder YrEngThst
(4) Put in loft insulation / extra loft insulation YrEngLoft
(5) Put in cavity wall insulation YrEngCvty
(6) Put in solid wall insulation YrEngSold
(7) Replace single glazed windows with double glazing YrEngGlaz
(8) Fit secondary glazing to windows YrEngScnd
(9) Put in solar water heating YrEngSwh
(10) Put in solar photovoltaic (PV) panels YrEngSpv
(11) None of these YrEngBNne

Figure 1: Demographic profile questions, EHS 2012-13

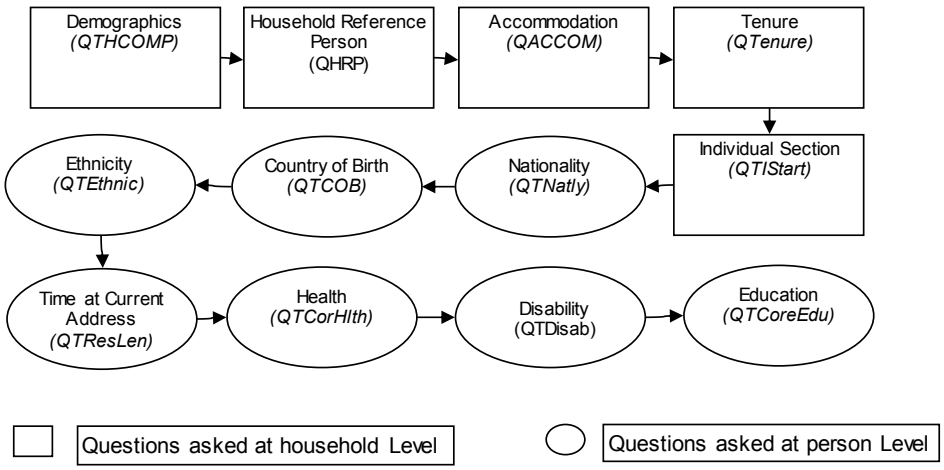
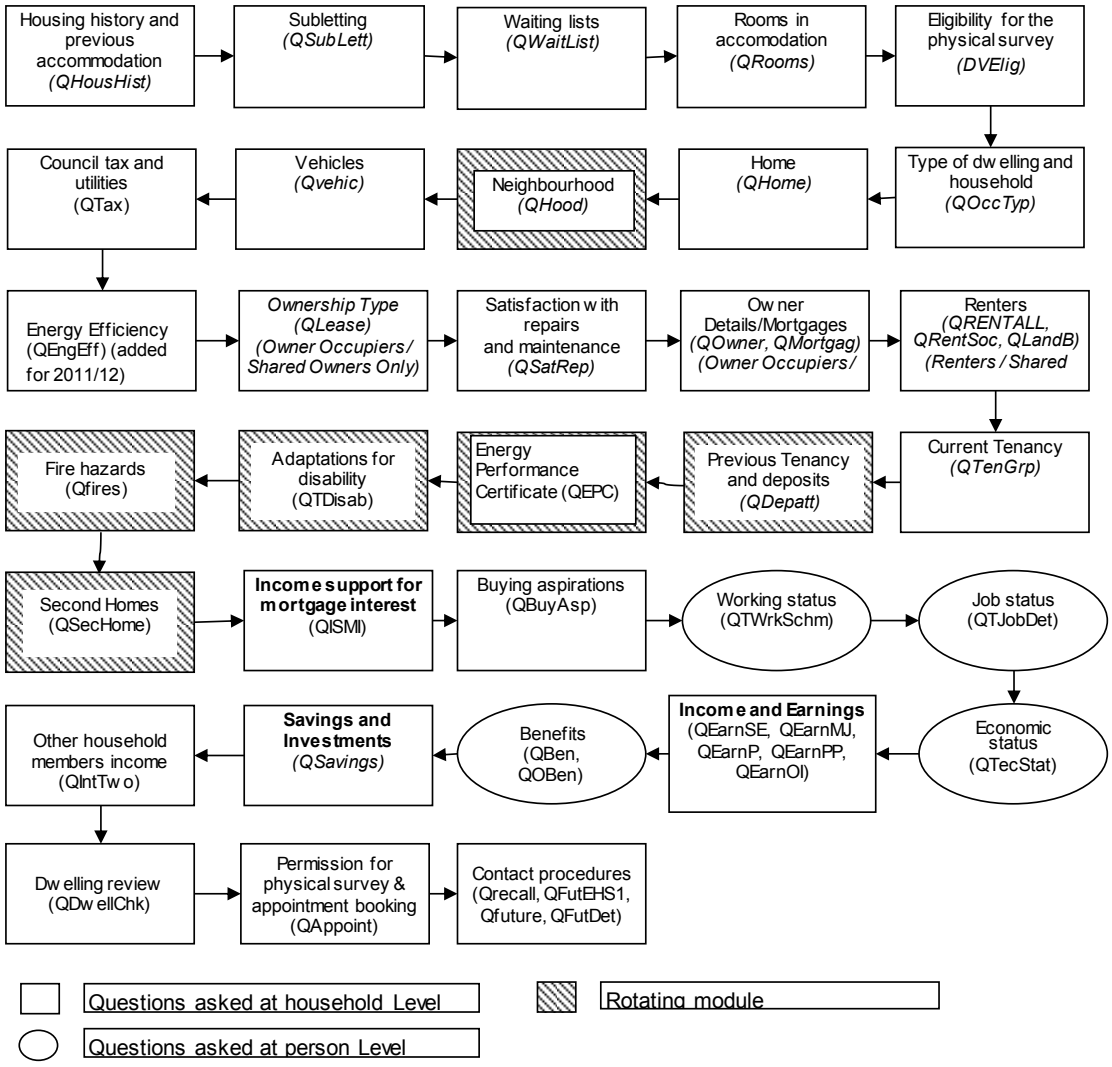


Figure 2: Housing questions, EHS 2012-13


Housing history and previous accommodation (QHousHist)
Subletting (QSubLett)
Waiting lists (QWaitList)
Rooms in accomodation (QRooms)
Eligibility for the physical survey (DVElig)
Type of dwelling and household (QOccTyp)
Home (QHome)
Neighbourhood (QHood)
Vehicles (Qvehic)
Council tax and utilities (QTax)
Energy Efficiency (QEngEff) (added for 2011/12)
Ownership Type (QLease) (Owner Occupiers/ Shared Owners Only)
Satisfaction with repairs and maintenance (QSatRep)
Owner Details/Mortgages (QOwner, QMortgag) (Owner Occupiers/
Renters (QRENTALL, QRentSoc, QLandB) (Renters / Shared
Current Tenancy (QTenGrp)
Previous Tenancy and deposits (QDepatt)
Energy Performance Certificate (QEPC)
Adaptations for disability (QTDisab)
Fire hazards (Qfires)
Second Homes (QSecHome)
Income support for mortgage interest (QISMI)
Buying aspirations (QBuyAsp)
Working status (QTWrkSchm)
Job status (QTJobDet)
Economic status (QTecStat)
Income and Earnings (QEarnSE, QEarnMJ, QEarnP, QEarnPP, QEarnOI)
Benefits (QBen, QOBen)
Savings and Investments (QSavings)
Other household members income (QIntTwo)
Dwelling review (QDwellChk)
Permission for physical survey & appointment booking (QAppoint)
Contact procedures (Qrecall, QFutEHS1, Qfuture, QFutDet)
Questions asked at household Level
Rotating module
Questions asked at person Level

How to match files

Matching files that exist at the household level

53 The key variable for matching at the dwelling/household level within or across any of the survey data is **aacode** variable (an eight character string variable). You must specify this as the key field when matching. All files on the database have been sorted by this variable so should match. 'Sort file' or 'split file' commands, however, re-sort the file. Matching with standard database files will fail if the data is not first re-sorted by aacode.

Matching lower level files (i.e. below dwelling or household level) up to the grossing file

54 Files at a lower level of organisation will have more than one case per household or dwelling. Some of them have exactly the same number of cases per address e.g. the windows.sav file has 7 cases for each address each representing a type of window, the chimneys.sav has 2 cases per address. With other files, the number of cases varies e.g. for people.sav the number of cases is the number of people in the household so one person households will just have one case, 6 person households will have 6 cases etc. As there is always more than one case per address, these cannot simply be matched into the dwelling or household level files to give results at the dwelling or household level. There are 2 ways to approach these files:

- Match the household or dwelling level files in as' tables'. Here the information at household or dwelling level will be copied for each person etc. within that aacode. The results with then represent the number of people, number of types of window etc. rather than the number of dwellings or number of households. This can be useful for very specific analyses, e.g. on numbers of people in non-decent dwellings.
- Aggregate the lower level file up to address code level and then match it in with the household or dwelling level file. (Most key variables for analysis have been derived at the household level so this will not normally be necessary).

How to apply grossing factors to obtain national estimates

55 The EHS household dataset is based on a stratified sample with over sampling of the rented tenures to achieve large enough sample sizes to produce reliable results for the social and private rented sectors.

Grossing factors have been calculated to:

- Compensate for the design of the sample i.e. the over sampling of some dwellings and under sampling of others; and
- Take account of non-response bias: the survey response rates achieved for different groups of households and dwellings

N.B results must be based on data weighted by the relevant grossing factor to produce national estimates.

56 The EHS housing stock dataset comes with its own grossing factors covering the dwelling sample of cases (see above). The grossing factors can be found in the file **general.sav**. The relevant variables are:

- Dwelling grossing – **aagpdxxyy**
- Household grossing – **aagphxxyy**

Where xxyy indicates the two years across which fieldwork was conducted. For example, **aagpd1112** is the dwelling grossing factor for the 2-year dataset covering EHS 2011-12 and EHS 2012-13 i.e. for the EHS 2012 dataset.

Aagpdxxyy should be used for any analysis in which the aim is to provide estimates of dwellings and that includes physical survey data (e.g. percentage of non- decent dwellings) while **aagphxxyy** should be used for any analysis in which the aim is to provide estimates of households and that includes interview survey data (e.g. percentage of households in non-decent dwellings).

The grossing factors can only be used on full 2-year datasets. They cannot be used on the data split into separate years.

57 The following number of cases and the grossed national totals are as follows:

Basic control totals for grossing dwellings and households in the housing stock datasets

Year	Dwellings - number of real cases	Dwellings - grossed counts (thousands)	Households - number of real cases	Households - grossed count (thousands)
2008	**16,150**	**22,239**	**15,523**	**21,530**
2009	**16,150**	**22,335**	**15,512**	**21,554**
2010	**16,670**	**22,386**	**16,047**	**21,600**
2011	**14,951**	**22,754**	**14,386**	**21,918**
2012	**12,763**	**22,718**	**12,269**	**21,935**

It is vital that these figures are used as a basic check on analysis i.e. that these grossed counts are being achieved before embarking on any more detailed analysis on sub-sections of households or dwellings.

How to deal with missing data

Missing data conventions

58 The DV files contain no system missing values. Where values are unknown or the question was not asked, specific codes are used and the values are set to user defined missing. The conventions adopted are:

a) Does not apply: -9 for single digit fields, -99 for 2-digit fields etc

b) Unknown: -8 for single digit fields, -88 for 2-digit fields etc

c) Not asked for a particular quarter: -2 for single digit fields, -22 for 2 digit fields etc

d) No partner: -7 for single digit fields, -77 for 2-digit fields etc

59 Users need to be aware of the implications of having missing values switched on or off within SPSS. It is useful for example to leave missing values switched on (the SPSS standard) if users wish to establish the level of missing data in a frequency or cross tab. The SPSS output Case Processing Summary will show the level of missing data. This can be used to decide on whether the level of missing data is significant and whether it is appropriate to re-distribute missing cases.

60 It is particularly important to leave missing values switched on when dealing with a continuous variable e.g. costs to make a home decent. Including -99, -88 etc as a normal value (i.e. not as a missing value) would distort any outputs.

When you create any new variables, you should declare missing values as below:

```
Missing values var1 (-9).
```

This code sets up '-9' as a user defined missing so will be excluded from any cross tabs or statistics of var1.

Users should note that the use of negative values for missing data is a change from the former EHCS

Dealing with missing data in analysis

61 Analysis that uses derived variables with no missing cases will have no problem with missing data. However, when undertaking analysis using other variables that have missing or unknown codes, these cases can be redistributed to produce national estimates.

62 Some procedures like 'cross tabulation' exclude missing cases from the table. To include these as the basis for any redistribution it is necessary to switch off the user defined missing values for each variable used. This is done by writing ***'missing values Var1 ().'*** into the SPSS syntax so that missing data will then be included in outputs.

63 Missing values can then be switched back on, by writing: **missing values var1 (-9)** into the SPSS syntax.

Checking if results are significant

64 An additional Excel spreadsheet, **significant change tester.xls**, is included to provide a quick method of checking whether changes in a particular indicator over time are statistically significant. There is an accompanying word document, which explains how the Excel macro works.

Appendix A: Content and level of interview survey files

File name	Contents	level
Adapt Hhld.sav	Information on adaptations made to a home.	household
Adapt Person.sav	Information on adaptations required for someone with a disability.	person
Adaptation Hhld.sav	Information on the adaptations a household currently has	household
Adaptation Person.sav	Information on adaptations needed for disability purposes	person
Attitudes.sav	Information on attitudes about current accommodation, area, and other opinion based questions	household
Contact.sav	Key contact information from the survey	household
Disability.sav	Information on types of disability	person
Dwelling.sav	Type of accommodation Age of accommodation, housing history Ownership type, buying aspirations	household
Employment.sav	Employment information	person
Energy.sav	Energy efficiency, Energy Performance Certificate.	household
Fire.sav	Fires in the accommodation	household
FirstImp_IS.sav	Key information on the first impressions of the dwelling and neighbourhood as recorded by interviewer	household
Hhldtype.sav	Tenure, type of dwelling and household, council tax and utilities.	household
Identity.sav	Country of birth, ethnicity and length of residence at current address	person
Income.sav	Information on amounts and sources of income including benefits	household
Owner.sav	Ownership details, mortgage	household
People.sav	Key info for each person (age, sex etc.) and relationships with other people in the household	person
Renter.sav	Renting, social renters, housing benefit, tenancy agreements and deposits,	household
Rooms.sav	Key information about the number of rooms and whether shared	household
Vacant.sav	Information relating to vacant properties from the doorstep form	household
Waitlist.sav	Waiting lists for social housing module	household

Appendix B: Content and level of physical survey files

File Name	Section on survey form	Level
Amenity.sav	5. Interior - Amenities	Dwelling
Around.sav	18. Exterior - plot of survey dwelling (not shared plots) 19. Around the house/module 20. Block 23. Local area	Dwelling
Chimney.sav	18. Exterior of house/module (front and rear views)	Type of chimney
Commac.sav	9. Common parts of flat surveyed: accessways (common parts by accessway, 3 cases per address)	Type of accessway
Common.sav	9. Common parts of flat surveyed	Dwelling
Damppc.sav	18. Exterior of house/module (front and rear views)	Type of damp proof course
Doors.sav	18. Exterior of survey dwelling (front and rear views)	Type of door
Dormers.sav	18. Exterior of house/module (front and rear views)	Type of dormer/bay
Elevate.sav	16. Elevation features 17. Specification of views	Dwelling
Firstimp_PS.sav	2. Dwelling identification 3. Dwelling description and occupancy 4. Module associated with address surveyed	Dwelling
Flatdets.sav	8. Details of flat	Dwelling
Hhsrs.sav	22. Housing Health and Safety Rating System	Dwelling
HQ.sav	7. Household questionnaire conducted by surveyor	Dwelling
Interior.sav	5. Interior 18. Exterior of house/module (conservatories)	Dwelling
Introoms.sav	5. Interior	Rooms
Numflats.sav	10. Number of flats in the module	Dwelling
Plotlvl.sav	18. Exterior - plot of survey dwelling (not shared plots)	Plot
Roofcov.sav	18. Exterior of house/module (front and rear views)	Type of roof covering
Rooffeat.sav	18. Exterior of house/module (front and rear views)	Type of roof feature
Roofstru.sav	18. Exterior of house/module (front and rear views)	Type of roof structure
Services.sav	5. Interior - Primary services 5. Interior - Space heating 6. Loft inspection	Dwelling
Shape.sav	12. House/module shape 13. External dimension of house/module 14. Material and construction of house/module 15. Improvements/alterations and when done	Dwelling
Shared.sav	11. Shared facilities and services	Dwelling
Structure.sav	21. Structural defects	Dwelling
Wallfin.sav	18. Exterior of house/module (front and rear views)	Type of wall finish
Wallstru.sav	18. Exterior of house/module (front and rear views)	Type of wall structure
Windows.sav	18. Exterior of survey dwelling (front and rear views)	Type of window

Appendix C: Content of detailed derived variables

Filename	**Variables**
Actual costs.sav	Variables for external and internal works
Dimensions.sav	Number of rooms, wall and floor areas
Energy performance.sav	SAP ratings, energy efficiency ratings, light, water and space heating costs, energy upgrade costs
Standardised costs.sav	All costs per square metre of floor area
HHSRS.sav	Variables based on 26 hazards
Equivalised Income.sav	Variables based on equivalised income